

GPT-4o judged scores across models – Case Study 1: China-sensitive topics

